



GENOME ANNOUNCEMENT

Open Access

Genomic characterization of a *Helicobacter pylori* isolate from a patient with gastric cancer in China

Yuanhai You¹, Lin Liu², Maojun Zhang¹, Yuanfang Zhu², Lihua He¹, Dongfang Li² and Jianzhong Zhang^{1*}

Abstract

Background: *Helicobacter pylori* is well known for its relationship with the occurrence of several severe gastric diseases. The mechanisms of pathogenesis triggered by *H. pylori* are less well known. In this study, we report the genome sequence and genomic characterizations of *H. pylori* strain HLJ039 that was isolated from a patient with gastric cancer in the Chinese province of Heilongjiang, where there is a high incidence of gastric cancer. To investigate potential genomic features that may be involved in pathogenesis of carcinoma, the genome was compared to three previously sequenced genomes in this area.

Result: We obtained 42 contigs with a total length of 1,611,192 bp and predicted 1,687 coding sequences. Compared to strains isolated from gastritis and ulcers in this area, 10 different regions were identified as being unique for HLJ039; they mainly encoded type II restriction-modification enzyme, type II m6A methylase, DNA-cytosine methyltransferase, DNA methylase, and hypothetical proteins. A unique 547-bp fragment sharing 93% identity with a hypothetical protein of *Helicobacter cinaedi* ATCC BAA-847 was not present in any other previous *H. pylori* strains. Phylogenetic analysis based on core genome single nucleotide polymorphisms shows that HLJ039 is defined as hspEAsia subgroup, which belongs to the hpEastAsia group.

Conclusion: DNA methylations, variations of the genomic regions involved in restriction and modification systems, are the "hot" regions that may be related to the mechanism of *H. pylori*-induced gastric cancer. The genome sequence will provide useful information for the deep mining of potential mechanisms related to East Asian gastric cancer.

Keywords: *Helicobacter pylori*, Gastric cancer, Next generation sequencing, Genomic features

Background

Helicobacter pylori, a Gram-negative bacterium that colonizes in the human stomach, has been widely recognized as a pathogenic bacteria related to the pathogenesis of gastritis, ulcers, and carcinoma [1-3]. The high genetic variability of *H. pylori* drives its dramatic ability to adapt to the gastric niche [4-9]. However, although many studies have been performed, its mechanisms are still not well elucidated.

With the rapid development of the next generation sequencing technology and reduced costs, it has become possible to perform large scale genome sequencing procedures

to obtain ample information about biological population structure and disease markers. Over the past few years, increasingly more *H. pylori* strains from different geographic regions, ethnicities, and diseases have been sequenced [10-12], and at least 50 genome sequences are currently available in public databases.

In a previous study, we published genome sequences of three strains recovered from patients with ulcers and atrophic gastritis in Heilongjiang province [13]. It is well known that *H. pylori* strains isolated from different geographic areas show dramatic genomic diversity [14]. Thus, at the genomic level, comparative analysis among strains with different clinical manifestations should initially eliminate such interference. Comparative genomic sequencing analysis of strains isolated from single patients could be a reliable way to eliminate such interference [15-17]. However, it is usually difficult to follow a

* Correspondence: zhangjianzhong@icdc.cn

¹Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, State Key Laboratory for Infectious Disease Prevention and Control and National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, People's Republic of China

Full list of author information is available at the end of the article

patient and obtain strains isolated from various unpredictable manifestations.

In this study, we reported a draft genome sequence of strain HLJ039 that was isolated from a patient with gastric cancer in Heilongjiang province. After integration with the other three genomes from the same area, initial comparative genomic analysis was performed to investigate the genetic features of gastric cancer isolates.

Methods

Strain selection

HLJ039 was isolated from an 84-year-old man with poorly differentiated stomach body cancer. Although some other gastric carcinoma-related *H. pylori* strains isolated from different areas, ethnicities, and populations in the world are present in public databases, we did not select these strains for our comparative analysis. The complex strain background will make it very difficult to identify reliable genomic characteristics that may be contributed to a specific disease like gastric cancer. As such, analyzing a specific geographic region, ethnicity, or population may be a more sensible way to find potential clues related to specific diseases. Therefore, in this study, we selected only three strains isolated from Heilongjiang province for the comparative analysis. These strains are very representative because Heilongjiang province has a high incidence of gastric diseases in China, especially for gastric cancer. In addition, the Chinese Heilongjiang province is near Korea and Japan. These east Asian countries reportedly have the highest incidence of gastric cancer worldwide [18,19].

Ethics approval

This research was approved by the meeting of ethics committee of national institute for communicable disease control and prevention, China CDC, according to Chinese ethics laws and regulations. NO:ICDC-2013001.

Genome sequencing and annotation

The strain was isolated from gastric mucosa and cultured on Columbia agar base supplemented with 5% sheep blood. DNA was extracted as previously described [20]. For each strain, whole-genome sequencing was performed using an Illumina Hiseq 2000 by generating paired-end libraries (500 bp and 2 kb) following the manufacturer's instructions. The read lengths were 90 bp and 50 bp for each library, from which more than 100 Mb of high-quality data was generated. The paired-end reads from the two libraries were de novo assembled into scaffolds using SOAPdenovo (<http://soap.genomics.org.cn>). Gene prediction was performed using Glimmer. The tRNA genes were searched for by tRNAscan-SE2, while the rRNA genes were searched for by RNAmmer3. Protein BLAST4 was

run using the translated coding sequences as a query against the reference sequence (*H. pylori* strain 51).

The genome was further annotated and functionally categorized by Rapid Annotation using Subsystem Technology (RAST). A subsystem is a set of functional roles that an annotator has decided are related. Subsystems frequently represent the collection of functional roles that compose a metabolic pathway, complex, or protein class [21].

Initial comparative genomic and phylogenetic analysis

To identify possible regions that may be involved in the pathogenesis of gastric cancer, MAUVE was used to compare HLJ039 with three additional isolates recovered from the same area [22]. As described previously, HLJ271 was recovered from a patient with gastric ulcer. HLJ193 and HLJ256 were recovered from patients with atrophic gastritis. Different regions (DRs) of HLJ039 were labeled along its chromosome location. DRs refer to coding sequence (CDS) insertion and deletion in HLJ039 compared to the other three genomes.

To define the phylogenetic characterization of HLJ039 using the publicly available *H. pylori* genome sequences, 53 whole genome sequences were extracted from GenBank for phylogenetic tree construction (Additional file 1). P12 was used as a reference genome. Comparisons were made using the nucmer program from MUMMER3 implemented in Panseq [23]. Genomes were fragmented into 500-bp segments that had to be present in all 54 genomes to be included in the core genome. Horizontally transferred genes usually have high genetic diversity among different strains, for example, the plasticity zones, which encode type IV secretion systems, R-M systems, or transferable genomic islands. According to the principle of multiple alignment by the use of Panseq, these potential horizontal genes would be removed from the core genes. Single nucleotide polymorphisms (SNPs) in the core genomes are determined and used to generate a Phylip-formatted file. Concatenated SNPs in length of 29,259-bp were used to construct a phylogenetic tree by using the neighbor-joining method in MEGA5. Bootstrap method was used to assess the stability of the phylogenetic relationships.

Genomic data deposition

This whole genome shotgun project has been deposited at DDBJ/EMBL/GenBank under accession number JAAA00000000, while version JAAA01000000 is described in this paper.

Quality assurance

The genomic DNA was extracted from a pure cultured *H. pylori* strain and confirmed using conventional biochemical tests (positive for urease, catalase, and oxidase).

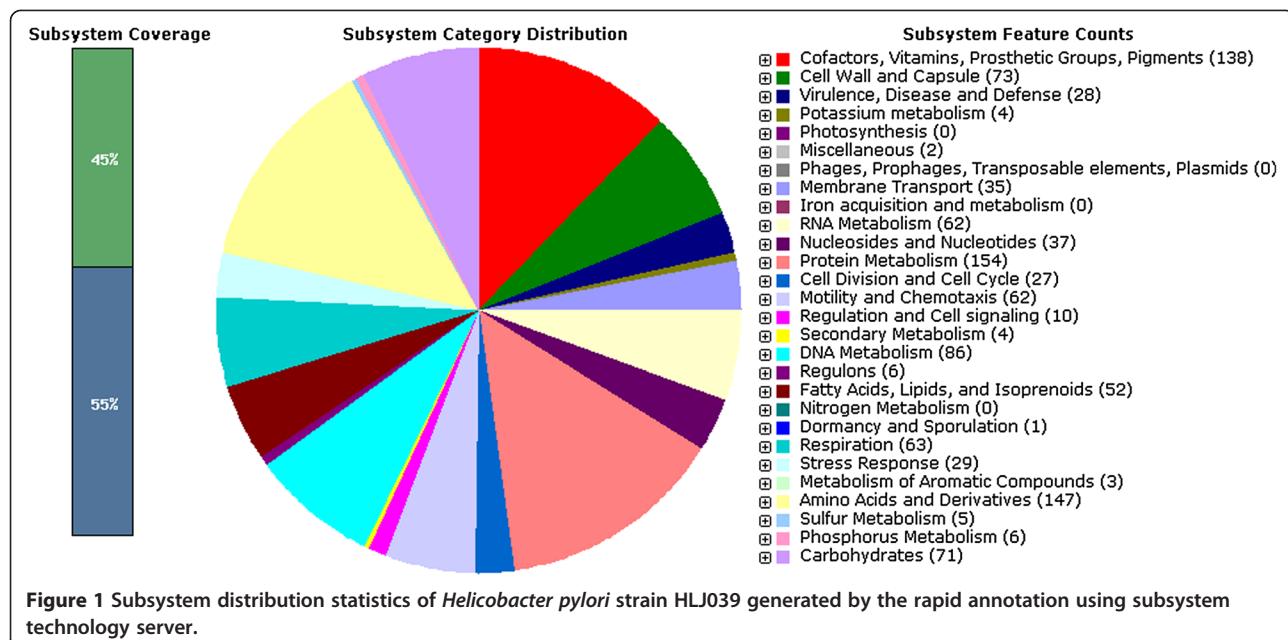


Figure 1 Subsystem distribution statistics of *Helicobacter pylori* strain HLJ039 generated by the rapid annotation using subsystem technology server.

The RAST server was used to evaluate potential heterogeneous contaminations.

Initial findings

We ultimately obtained 42 contigs with a total length of 1,611,192 bp and predicted 1,687 CDS within the draft genome of strain HLJ039. Additional information is included in the sequencing reports of HLJ039 (Additional file 2). The G + C content was 38.72%. The subsystem distribution and general information about the potential functional distribution of HLJ039 are shown in Figure 1. Compared to the additional three HLJ genomes, HLJ039 has 10 different regions (DRs). Detailed information about these fragments is shown in Table 1. The locations of these DRs are labeled in the whole genome (Figure 2). Approximately half of these sequences encoded hypothetical proteins. Most of the DR sequences encoded proteins involved in DNA methylase and a restriction modification enzyme. Notably, a unique 547-bp fragment (DR9) sharing 93% identity with a hypothetical protein of *Helicobacter cinaedi* ATCC BAA-847 was found that had never been present in any other *H. pylori* strains previously, which indicated a possible horizontal gene transfer between *H. pylori* and *H. cinaedi*. DR9, located in scaffold 5, inserts into a 1,371-bp gene encoding type III restriction endonuclease, which is responsible for adenine-specific DNA methylase modifications.

All of the above findings highlight the important role of DNA restriction modification systems in *H. pylori* genomic recombination. A total of 29,259 core SNPs were found among the 54 analyzed genome sequences. Based

on a core genome SNP analysis of 54 *H. pylori* strains distributed in various worldwide regions, a phylogenetic tree was generated to show the HLJ039 subtype. All strains were classified into different groups defined by earlier studies according to multilocus sequence typing [24,25]. Figure 3 shows that HLJ039 was defined as belonging to the hspEAsia subgroup, which belonged to the hpEastAsia group.

Table 1 Basic information of the different regions (DRs) in HLJ039

DR	Start	End	Gene description
DR1	145736	180926	25 hypothetical proteins, VirB4, DNA topoisomerase I, ParA, Mobile element protein, First ORF in transposon ISC1904
DR2	618752	619703	Fucosyltransferase
DR3	740131	740654	Hypothetical protein
DR4	1200420	1202309	Hypothetical protein
			DNA-cytosine methyltransferase
DR5	1254233	1256053	Hypothetical protein
DR6	1335551	1337398	Type II m6A methylase (hinFIM) hypAIVR
DR7	1393932	1394805	Hypothetical protein
DR8	1443251	1445196	Type II DNA modification enzyme
			hypothetical protein
DR9	1484058	1484604	Hypothetical protein sharing 93% identity with a fragment of <i>Helicobacter cinaedi</i> ATCC BAA-847
DR10	1538060	1539662	Type IIIG restriction and modification enzyme

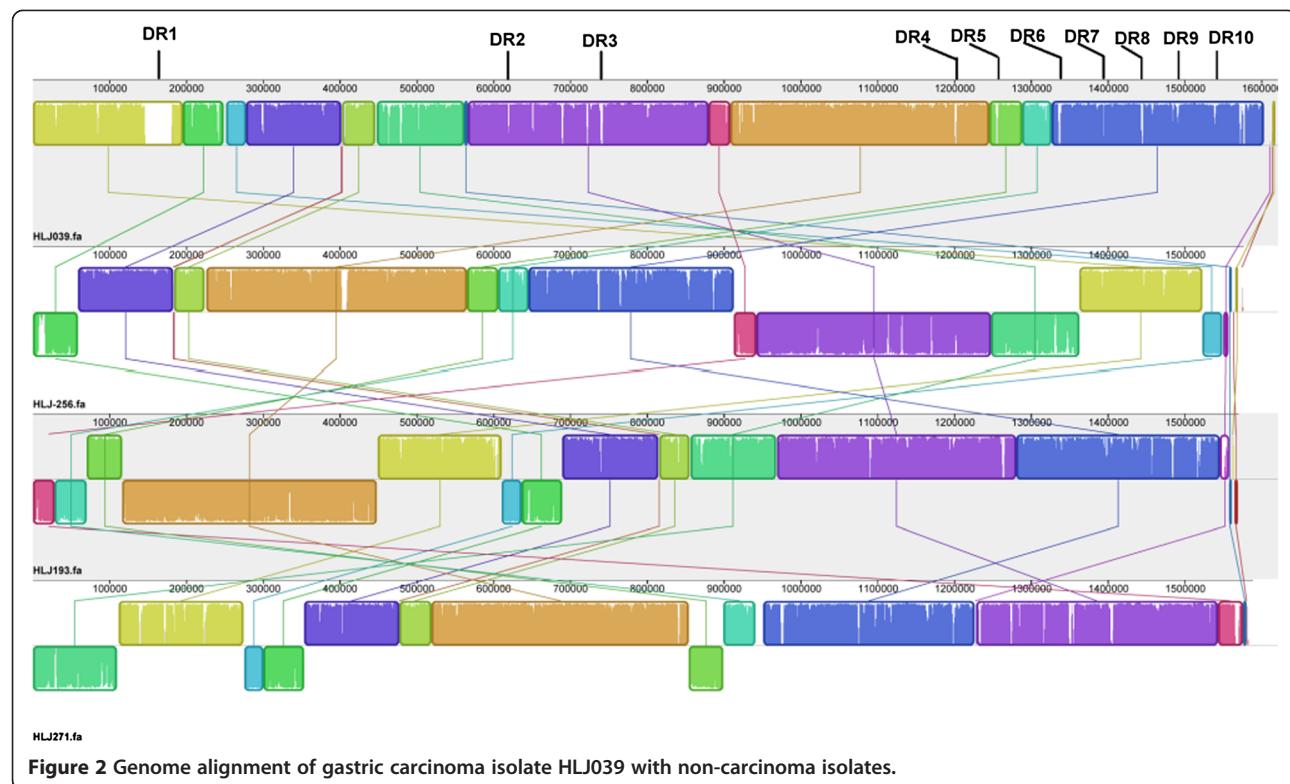


Figure 2 Genome alignment of gastric carcinoma isolate HLJ039 with non-carcinoma isolates.

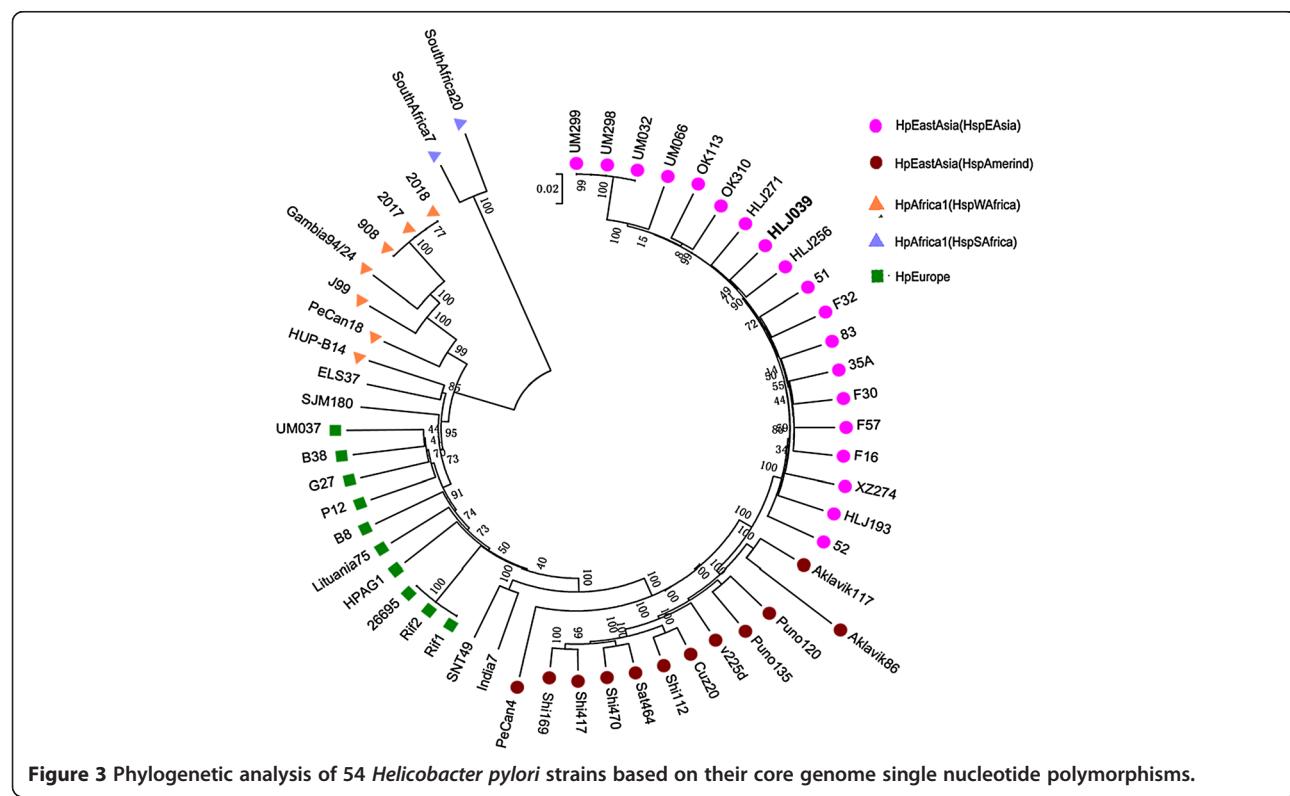


Figure 3 Phylogenetic analysis of 54 *Helicobacter pylori* strains based on their core genome single nucleotide polymorphisms.

Note: Different regions (DRs) refers to coding sequence insertion and deletion in HLJ039 compared to the other three genomes.

Future directions

The incidence of gastric carcinoma in East Asian countries is quite high [18,19]. To explore the potential pathogenic mechanisms that may contribute to this phenomenon, more East Asian *H. pylori* strains must first be sequenced. The strains selected for sequencing should be representative and eliminate geographic variation. Our future directions will focus on large-scale genomic sequencing of different clinical isolates from areas with a high incidence of gastric cancer. More detailed analyses involved in DNA methylation as well as restriction and modification systems would be the most attractive directions for studies of *H. pylori*-induced gastric cancer.

Consent

Written informed consent was obtained from the patient for the publication of this report and any accompanying images.

Availability of supporting data

Additional data supporting the results reported here are included within the additional files.

Additional files

Additional file 1: General information for the publicly available genomes.

Additional file 2: Assembly information for HLJ039.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YY performed the bioinformatics analysis and wrote the manuscript; MZ and LH were responsible for bacteria isolation and identification; LL, XH and YZ performed genomic sequencing; JZ and PN designed the study and provided financial support for this work. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by a fund for China Mega-Project for Infectious Disease (2011ZX10004-001) and a grant from the National Technology R&D Program in the 12th Five-Year Plan of China (2012BAI06B02).

Author details

¹Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, State Key Laboratory for Infectious Disease Prevention and Control and National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, People's Republic of China. ²BGI-Shenzhen, Shenzhen, People's Republic of China.

Received: 2 January 2014 Accepted: 18 February 2014

Published: 24 February 2014

References

1. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N, Schlemper RJ: *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med* 2001, **345**:784-789.
2. Marshall B: *Helicobacter pylori*. *Am J Gastroenterol* 1994, **89**:S116-S128.
3. Gerhard M, Rad R, Prinz C, Naumann M: Pathogenesis of *Helicobacter pylori* infection. *Helicobacter* 2002, **7**(Suppl 1):17-23.
4. Ahmed N: Replicative genomics can help *Helicobacter* fraternity usher in good times. *Gut Pathog* 2010, **2**:25.
5. Falush D, Kraft C, Taylor NS, Correa P, Fox JG, Achtman M, Suerbaum S: Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc Natl Acad Sci USA* 2001, **98**:15056-15061.
6. Gressmann H, Linz B, Ghai R, Pleissner KP, Schlapbach R, Yamaoka Y, Kraft C, Suerbaum S, Meyer TF, Achtman M: Gain and loss of multiple genes during the evolution of *Helicobacter pylori*. *PLoS Genet* 2005, **1**:e43.
7. Ahmed N, Dobrindt U, Hacker J, Hasnain SE: Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. *Nat Rev Microbiol* 2008, **6**:387-394.
8. Ahmed N: A flood of microbial genomes—do we need more? *PLoS One* 2009, **4**:e5831.
9. Ahmed N, Tenguria S, Nandanwar N: *Helicobacter pylori*-a seasoned pathogen by any other name. *Gut Pathog* 2009, **1**:24.
10. Ahmed N, Loke MF, Kumar N, Vadivelu J: *Helicobacter pylori* in 2013: multiplying genomes, emerging insights. *Helicobacter* 2013, **18**(Suppl 1):1-4.
11. Lu W, Wise MJ, Tay CY, Windsor HM, Marshall BJ, Peacock C, Perkins T: Comparative analysis of the full genome of *Helicobacter pylori* isolate Sahul64 identifies genes of high divergence. *J Bacteriol* 2014, **196**(5):1073-1083.
12. Kumar N, Mukhopadhyay AK, Patra R, De R, Baddam R, Shaik S, Alam J, Tiruvayipati S, Ahmed N: Next-generation sequencing and de novo assembly, genome organization, and comparative genomic analyses of the genomes of two *Helicobacter pylori* isolates from duodenal ulcer patients in India. *J Bacteriol* 2012, **194**(21):5963-5964.
13. Yuanhai Y, Lin L, Maojun Z, Xifang H, Lihua H, Yuanfang Z, Peixiang N, Jianzhong Z: Genome sequences of three *Helicobacter pylori* strains isolated from atrophic gastritis and gastric ulcer patients in China. *J Bacteriol* 2012, **194**(22):6314-6315.
14. Linz B, Schuster SC: Genomic diversity in *Helicobacter* and related organisms. *Res Microbiol* 2007, **158**:737-744.
15. Avasthi TS, Devi SH, Taylor TD, Kumar N, Baddam R, Kondo S, Suzuki Y, Lamouliatte H, Mégraud F, Ahmed N: Genomes of Two chronological isolates (*Helicobacter pylori* 2017 and 2018) of the West African *Helicobacter pylori* strain 908 obtained from a single patient. *J Bacteriol* 2011, **193**(13):3385-3386.
16. Gustavsson A, Unemo M, Blomberg B, Danielsson D: Genotypic and phenotypic stability of *Helicobacter pylori* markers in a nine-year follow-up study of patients with noneradicated infection. *Dig Dis Sci* 2005, **50**:375-380.
17. Israel DA, Salama N, Krishna U, Rieger UM, Atherton JC, Falkow S, Peek RM Jr: *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci USA* 2001, **98**:14625-14630.
18. Stewart BW, Kleihues P: *World Cancer Report*. Lyon: IARC Press; 2003.
19. Crew KD, Neugut AI: Epidemiology of gastric cancer. *World J Gastroenterol* 2006, **12**(3):354-362.
20. Yuanhai Y, Lihua H, Maojun Z, Jianying F, Yixin G, Binghua Z, Xiaoxia T, Jianzhong Z: Comparative genomics of *Helicobacter pylori* strains of China associated with different clinical outcome. *PLoS ONE* 2012, **7**(6):e38528.
21. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsmma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O: The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008, **9**:75.
22. Darling AC, Mau B, Blattner FR, Perna NT: Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 2004, **14**(7):1394-1403.
23. Laing C, Buchanan C, Taboada EN, Zhang Y, Kropinski A, Villegas A, Thomas JE, Gannon VP: Pan-genome sequence analysis using Panseq: an online tool

for the rapid analysis of core and accessory genomic regions. *BMC Bioinforma* 2010, **11**:461.

24. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, Blaser MJ, Graham DY, Vacher S, Perez-Perez GI, Yamaoka Y, Mégraud F, Otto K, Reichard U, Katzowitsch E, Wang X, Achtman M, Suerbaum S: **Traces of human migrations in *Helicobacter pylori* populations.** *Science* 2003, **299**:1582–1585.

25. Suzuki R, Shiota S, Yamaoka Y: **Molecular epidemiology, population genetics, and pathogenic role of *Helicobacter pylori*.** *Infect Genet Evol* 2012, **12**(2):203–213.

doi:10.1186/1757-4749-6-5

Cite this article as: You et al.: Genomic characterization of a *Helicobacter pylori* isolate from a patient with gastric cancer in China. *Gut Pathogens* 2014 **6**:5.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

